

Original citation:

Clayton, Matthew and Halliday, Daniel. (2017) Big data and the liberal conception of education. Theory and Research in Education.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/92353>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

This had been posted ahead of publication.

Clayton, Matthew and Halliday, Daniel. (2017) Big data and the liberal conception of education. Theory and Research in Education. Copyright © 2017 The Authors. Reprinted by permission of SAGE Publications.

<http://journals.sagepub.com/home/tre>

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Big Data and the Liberal Conception of Education

Matthew Clayton
Warwick University, United Kingdom

Daniel Halliday
University of Melbourne, Australia

ABSTRACT: This paper develops a perspective on big data in education, drawing on a broadly liberal conception of education's primary purpose. We focus especially on the rise of so-called learning analytics, and the associated rise of digitization, which we evaluate according to the liberal view that education should seek to cultivate individuality, and proceed partly by way of experimentation and with an emphasis on civic education. Our argument is not that the use of big data is wholly out of place in education. Indeed, it might have significant value in pursuit of certain educational aims. Nevertheless, the liberal conception shows how education is distinct from other domains in which big data is being applied, in ways that suggest that considerable caution must be exercised when they are used in educational contexts.

KEYWORDS: Liberalism; Individuality; Big Data; Civic Education

1. Individuality and Liberalism

In the autobiography of John Stuart Mill, one encounters the following opinion about education:

Most boys or youths who have had much knowledge drilled into them, have their mental capacities not strengthened, but over-laid by it. They are crammed with mere facts, and with the opinions or phrases of other people, and these are accepted as a substitute for the power to form opinions of their own: and thus the sons of eminent fathers, who have spared no pains in their education, so often grow up mere parroters of what they have learnt, incapable of using their minds except in the furrows traced for them. (1989a: 44 – 45)

These remarks, though strongly put, reflect a relatively sophisticated and attractive view about what the education of children is supposed to achieve and how easy it is for institutions to miss the mark, particularly when training the elite.¹ Mill believed that education ought to do at least two things. First, education should cultivate individuality in those who receive it. Pursuit of this goal counts against attempting to realize a uniform, prior conception of what makes for an educated individual. Second, education should be experimental, in the sense that, within certain constraints, its providers should be free and encouraged to develop different and competing curricular, rather than have a single method foisted on them by any higher authority. A third goal is that education should train young people for citizenship. This goal of 'civic education' includes an emphasis, in education, on training children to appreciate rule of law and respect demographic diversity (understood as involving a variety of conceptions of the good within society). Though Mill was somewhat sceptical about government's role in education², the

¹ Such concerns continue to endure. For a much for detailed discussion of where things are up to in the United States, see Deresiewicz (2015).

² In the closing chapter of *On Liberty*, Mill remarked that "a general State education is a mere contrivance for moulding people to be exactly like one another" (1989b: 106). Mill's reasons for making this claim

pursuit of civic education is compatible, even extends, his views about individuality especially. Taken together, the joint pursuit of these goals requires an expectation that students will turn out quite differently from each other, and that their schools might be allowed to remain quite different from each other as well. At the very least, something has gone wrong when students are churned out by an education system that resembles an industrial production line in the homogeneity of what it's producing.

In this paper, we will use these ideas, which we consider central to any liberal conception of education, to develop a perspective on the use of big data in education. A quick way of describing our position is that education is, in two chief respects, quite unlike other contexts in which the use of big data is becoming prominent and where its use may be considered successful. This is because whereas many other goods and services are produced in an invariant form and consumed at the user's discretion, education is (or ought to be) produced in a more tailored form and is delivered coercively. An example of a good that typically satisfies the first description is healthcare. Broadly speaking, the quality of health outcomes is measured according to criteria that are more or less invariant across users of healthcare services.³ Although humans differ in important ways, we share a wide range of more or less identical physiological needs and vulnerabilities. Any measure of the quality of healthcare must reflect this fact. It is plausible to assume that some burdensome condition, such as high blood pressure, is a bad thing regardless of who the patient is. Though there might be differences of degree, high blood pressure is not the sort of thing that is considered bad for some patients and good, or neutral, for others. But the healthcare industry is set up so that its services are consumed according to users' discretion. Patients are not simply *forced* to take whatever treatment is known to alleviate high blood pressure, or anything else. When some patient is vulnerable to the onset of such a condition, and (say) must take certain drugs to reduce this risk, medical institutions do not proceed by forcing them to do so. In the normal course of events, a patient is simply advised about the situation, given a prescription, and left free to take or not take the drugs that will help them. These points capture the fact that healthcare provision is based on homogeneity among users and non-coercive delivery.

With education, at least in the K-12 range, things are the other way round. There is no single, settled, very precise view about what children should be getting out of education that is on a par, for example, with what patients with high blood pressure should be getting by way of health outcomes. And whatever education is ultimately provided for children, it is delivered in a largely coercive manner. This is true both in the sense that children are legally required to attend school in the first place, will generally have to accept what choices their parents make as to which school this is, and will generally be told what to do by their teachers in the course of the typical school day. We are not suggesting that there is anything morally objectionable about all this. Apart from anything else, coercion need not be morally objectionable in the first place,

(and some closely related ones) deserve greater discussion than we can provide here, though we think his pessimism, while not altogether implausible, somewhat overstated.

³ Whether education and healthcare have homogenous or heterogeneous goals may vary according to the level of abstraction with which their goals are characterized. The goals of medical treatment will start to come apart for different individuals in many cases, e.g. a desired body mass index for different individuals. But we use this analogy for illustrative purposes: Our arguments below do not strictly depend on any view about whether healthcare is as different from education as we're suggesting it is. At least, the liberal conception of education does not depend on education's being different from healthcare or other domains of institutional provision, but rather from the independent considerations about what education is for, as laid out here.

and the paternalistic treatment of children has always been one of the less controversial contexts in which coercion might operate. And there are all sorts of reasons for requiring children to attend school, not least the fact that even the best family in the world cannot provide everything that a child can get from even a moderately good school (particularly when it comes to social integration). But the coercive nature of education, though justifiable, needs to be constrained by part of what motivates formal education in the first place, which is its role in helping children to prepare for adult life and, to some extent, enrich their childhoods. Education needs to create space for children to develop in their own way, not just force them through some planned agenda. This is not simply a constraint on how to educate children, but a positive view about what it really means for children to gain an education most worth receiving. This is what exercised Mill when he wrote the remarks we quoted. We think a strong case can be made for this sort of view, and that the evaluation of big data needs to proceed accordingly.

Our goal is not to be dismissive of big data in education. Indeed, towards the end of our reflections we note the valuable role they can play in pursuit of particular educational aims, such as the maintenance of academic integrity. Nevertheless, we are concerned to highlight what is distinctive about education compared to other sectors in which important services are provided. We want to argue that big data's application to education probably needs to be constrained, and that there is at least good reason to be wary of any naïve extension of various prior applications in other domains. This just reflects the fact that while it is plausible to hold a liberal conception of education, there is no strongly analogous liberal conception of healthcare. In short, the worry is that the application of big data tends to presuppose a certain homogeneity among the parties impacted by any such application. Although the articulation of this worry is our main goal here, we will also make some tentative but positive suggestions as to where the use of educational big data is compatible with, and may even advance, the liberal conception of education.

Section two provides a brief and largely uncontroversial summary of why individuality and experimenting are important educational goals. In section three, we'll develop some concerns about big data in education, particularly the category of 'learning analytics' and the increased use of digitization. Section four raises some more positive suggestions about the place of big data in education, and concludes.

2. The goals of liberal education – a brief summary

Broadly speaking, the central point of the liberal conception of education is that a just educational system effectively prepares children for life as autonomous adults.⁴ This has implications for both academic and civic aspects of education. Precisely what the liberal conception recommends is very much up for debate, and we will work with a fairly formal understanding of what it says, so as to avoid getting bogged down in internal disagreements. What is most important is the idea that, in order to train autonomous adults, educational

⁴ Here we follow Brighouse (2000; 2006). Other important defences of the liberal view are Feinberg (1992), Callan (1997), Gutmann (1999), Levinson (1999), Reich (2002), Clayton (2006), and MacMullen (2007). We should say that we use 'liberal' in a relatively broad sense that emphasizes the importance of individuality and autonomy. Our endorsement of these values may be compatible with a variety of perspectives in moral and political philosophy that do not necessarily identify as fundamentally liberal, such as a utilitarian outlook.

institutions need to cultivate individuality and freedom, and avoid homogeneity in the young adults that they release into the world when their schooling is finished.

The liberal conception of education has the advantage not just of providing foundational guidance for how schooling should proceed, but of explaining why the very existence of institutionalized education is morally justified. The liberal conception draws much of its strength from the fact that the family, no matter how idealized, simply cannot provide all of the goods that help a child prepare for life as an autonomous adult. For one thing, families cannot expose a child to a variety of conceptions of the good, or to peers from a wide range of demographics. This is true even for families that can absorb the material costs of providing such benefits to their children (which, of course, most families cannot). A school environment, suitably integrated, can do much better. This is ultimately due to the fact that effective preparation for adult life is greatly enhanced when a child is placed in an information-rich environment in which they are exposed to a wide range of stimuli. It should be emphasized that things are not so simple as to allow that children can be put in an environment together and simply left to integrate in ways that are guaranteed to be mutually beneficial and free from hazards.⁵ But social integration is an important goal of liberal education.⁶ There are a host of questions about how such environments should be managed. The point is that, by emphasizing the way in which schooling is necessary to expose children to a range of inputs, the liberal conception helps justify why education should exist in an institutionalized (as opposed to, say, family-operated) manner in the first place.

As well as justifying the existence of schooling, the liberal conception has implications for how education should be delivered within educational institutions. To put the point crudely, children develop in different ways, and they develop in different directions. And this is a good thing. The former point reflects the truism that what works for one student, in terms of the facilitation of academic development, might not work as well, or as quickly, for other students. The latter point has to do with the way in which children, as users of education, are perhaps most different from (say) cancer patients. While we generally want to secure the same outcome for patients (remission of cancerous cells), education can, and should, create different and complementary adults. This draws on the old liberal insight that a society will be more prosperous, both economically and otherwise, when its citizens display a diversity of skills and personalities. This thought goes back at least as far as Adam Smith's remark that, where humans are concerned, 'the most dissimilar geniuses are of use to one another' (1993: 11). Smith was partly making a point about economic production and the division of labour into specialized tasks so that labour could develop in ways that resulted in superior levels of production. As he put it, a diverse set of talents enhances 'the common stock' with which humans might produce things for each other. If an education system produces a largely homogenous body of rote-trained students (the sort of result that Mill disparaged), then it will probably fail to supply the economy with very good human capital – something that is, of course, partly a measure of what sort of labour is available to serve as inputs into economic production. Liberals may disagree among themselves about the importance of economic growth, and the extent to which its importance should shape the delivery of education. But economic production aside, there is also something to be said for the sheer (but complementary) fact that human beings differ from each

⁵ For a recent discussion of this problem from a philosophical perspective, see MacFarlane (forthcoming).

⁶ For more on how the pursuit of social integration should shape justice in education, see Anderson (2007) and Satz (2007).

other in terms of what they want out of life. Insofar as children start to work out what sort of life they want to pursue before they leave school, our schools should give them some opportunity to start heading off in different directions. This is not simply a point about the role of diverse human capital in enabling greater economic production, nor is it obvious that Smith himself was intending to say anything so narrow.

Human diversity may be desirable for other reasons. A society of varied individuals will have its share of disagreement, perhaps even conflict, about the good life. We follow John Rawls in noting that diversity about the good life is a natural outgrowth of a democratic culture in which people enjoy various basic freedoms.⁷ The right response to this is not to try to prevent it, even if one at times finds reason to regret it. Instead, we need to facilitate an environment in which such diversity can come about without antagonism and intolerance, in which people can feel comfortable about their own conception of the good and comfortable about the presence of others. Again, this is helped as much by social integration as by any particular plan of learning or curriculum. School is often the first place in which children become exposed to people who are very different from themselves and the members of their family. It is also a place in which they can begin to learn effectively about what it means to participate in a democratic society, along with other aspects of civic education.

Setting aside the history of political philosophy, we might wonder how individuality should best be cultivated in a contemporary educational setting. The short (and negative) answer is that it cannot be easily cultivated by the one-dimensional pursuit of planned educational outcomes. This rests partly on the liberal contention that a person's ends should not be set by others who happen to occupy a position of coercive authority. Instead, those equipped with such authority (schools included) can facilitate an individual's independent development of ends by providing a facilitative environment while holding back from any aggressive attempt to shape how things turn out. In this respect, liberal education bears some resemblance to a plausible ethics of parenting.⁸ We tend to think poorly of parents who seek to mould their children according to a very strong and pre-conceived idea of how they should turn out. But this is in spite of our equally strong conviction that parents ought to provide substantial levels of support for their child's development. The compatibility of these views is explained by the fact that it is possible to provide support in ways largely free from planning whatever outcomes such support might facilitate. The liberal conception of education seeks to emphasize a similar compatibility between state support for (even laws requiring) the schooling of children, without a particularly strong set of ideas about how children should turn out once educated. One important disanalogy between parents and the state is that duties of support may be much stronger when held by educational institutions than when held by individual parents. One view is that parents have some right to shape children according to their own values, as parent-child relationships are plausibly opportunities for both parties to advance their wellbeing. This grants parents some degree of latitude to engage in what we might call 'planning', at least when it comes to shaping the values of their children.⁹ Schools, on the other hand, do not have a level of

⁷ See especially Rawls's remarks about what he called 'the fact of reasonable pluralism' (1993: 36).

⁸ The remarks in this paragraph draw on the account of parenting developed in Clayton (forthcoming).

⁹ For philosophical discussion of this view, see Brighouse & Swift (2014: Ch.6). It is possible to deny the claim that parents really have a distinctive right to shape children according to their own values, and one of us does indeed deny this claim (Clayton forthcoming). But we both endorse the weaker view that there is a significant difference between the costs that schools can more easily be expected to absorb the costs of exposing children to a wide range of conceptions of the good.

wellbeing that is at stake when it comes to how they influence the children that attend them. It is plausible, then, to expect schools and other institutions to absorb more of the costs of contributing to the development of children, compared to the costs that parents might be expected to absorb.

The liberal conception is not exhaustive or free from limits. It recognises that there is a need for some basic skills, such as literacy and numeracy, which (so far as is possible) all students ought to be taught. There is also some pressure for education to enable the growth vocational skills for the labour market. These, again, may aim somewhat at producing a certain ‘type’ of student so as to achieve desirable levels of standardization, at least so that employers know what they are getting when they hire recent school leavers or university graduates. It is hard, for example, to make a very strong case for experimentation and individuality when it comes to the style of teaching in medical schools and for the training of electricians. There is a lot to be said for having a medical professional whose work is ultimately quite standardized – nobody really wants a doctor who likes to experiment (or a pilot or plumber). Again, though, the pursuit of these goals should show some sensitivity to the value of individuality – not all students will gain basic skills in the same way, and there is a case for making vocational training, which may occur later in a student’s educational career, more an option than a requirement. A high school curriculum that foists vocational training on teenage students, as happens in some countries, may be harder to justify than one that gives them more freedom to choose what they are taught.

Society needs education for a variety of reasons besides the cultivation of diverse autonomous adults. This will inevitably lead to trade-offs when it comes to designing schools and their curricula. Students need space to develop their own conception of the good, integrate with a diverse group of peers, and make somewhat independent choices as to what academic and vocational paths they might begin to head down. But some standardization might be necessary and justified. The point to stress is just that the balance needs to be maintained rather than upset. The worry we have about big data is that, absent proper care, it threatens to suppress the distinctively liberal goals of cultivating diversity and individuality through independence and experimentation, in favour of an undue degree of standardization. The plausibility of this worry will, of course, become more clear when we look at the way in which educational big data can be put to work.

3. Some concerns about big data

Broadly speaking, ‘big data’ refers to bodies of information that are large enough to be impenetrable to the cognitive processing powers of any human being. The term is sometimes used to refer not just to the sheer size of datasets themselves, but also to their internal complexity and/or the associated complexity of whatever computational process is used to analyse them.¹⁰ Typically, the application of big data consists in the use of algorithms that are designed to process it in ways that involve a certain recommendation or decision to be made. The ideas of data and algorithms need to be kept distinct, but need to be taken together if the significance of big data is to be appreciated. Worries about big data are very often worries about the combined effect of data that serves as inputs to a decision procedure carried out by an algorithm.

¹⁰ For more on the conceptualization of the term ‘big data’, see the discussion in Mittelstadt & Floridi (2015: 309-311).

The use of data has the potential to enable great improvements in modelling and predicting the behaviour of its subject matter. There are already many examples of where big data has had a big impact. We have already mentioned healthcare, but other industries have demonstrated comparably impressive progress. Big data is now used to power spam filters in email programs, pricing algorithms in retail and service industries, and targeted advertising. There is every reason to think that these applications have brought about substantial benefits to the users of these systems. Those of us who are old enough to have been using the internet for more than ten years can probably attest to the fact that less junk mail gets into our inboxes than in the old days, and how much easier it is to book flights and hotels from a wide range of options.¹¹

There are some very general concerns about big data that apply quite widely, that is, across various industries or institutions in which big data has gained, or may soon gain, an influence. Some worry that data will get so 'big' that humans no longer have the ability to oversee whatever systems are processing it. If something starts going wrong with the system, it may be that nobody is in a position to notice until things have got quite advanced and perhaps irreversible. The global financial crisis is sometimes mentioned as an example of this sort of case – financial products, and the contracts associated with their sale, had become far too complex for humans to make sense of. Another concern is that, insofar as big data can be owned and withheld from some that might want to use it, it may place too much power in the hands of those able to gather it. In addition, there is a serious problem of causal feedback loops, where the use of an algorithm to process data will eventually shape the kind of data that the same algorithm continues to process later on, again perhaps without anyone noticing. These are very general worries and it is worth stressing that they certainly extend to the educational context.¹² But, since they are not particularly germane to the liberal conception of education, we shall leave the greater part of their further investigation to others.

Instead, we wish to say something more specific to the sort of big data that is being used, or expected to be used, within the educational context. One general point that we do want to emphasize is that we should be wary of the institutional structure in which any use of big data is embedded. The reliability of any data (big or small) may be undermined given certain factors in the institutional background. In the case of education, data on student performance will be unreliable, for example, if teachers have been given (perverse) incentives to cheat when grading student exams. Continued efforts to allow such data to guide important decisions about how teachers and students are treated is likely to increase the damage that might be done when such background factors are present. We need to keep this fact in mind when considering big data.

Another general observation about the assessment of big data in education is that in a complete account it would be important to evaluate it in the context of the available alternatives. The use of such data might constitute significant progress in an educational environment in which teachers are underpaid, undertrained and undervalued, and, consequently, tend to teach in rote ways that are not congenial to the liberal model we have outlined. In a different educational context—one that boasts an able, highly trained, independent-minded teaching force, for example—the use of big data to enhance learning

¹¹ For some more examples, and general background about the rise of big data, see Mayer-Schoenberger & Cukier (2013). For an important articulation of some worries about big data and algorithms, see O'Neill (2016).

¹² See especially Slade & Prinsloo (2013).

might, as we shall argue, be suboptimal, because teacher-led enhancement would do a better job. In the remarks, below we focus on possible problems with the use of big data in education without addressing the thorny issue of comparing different kinds of educational regime in our non-ideal world.

One type of educational data that is frequently reported fits under the category of ‘learning analytics’.¹³ Broadly speaking, this has to do with a dramatic increase to record certain things that students do when they learn, and detect certain patterns and correlations. Students increasingly work with computers, through which they leave so-called digital trails that can be stored indefinitely and shared around on a global scale. Old fashioned modes of education, which are not computer based - think of vocal classroom teaching and students reading printed books – do not leave such trails. The new digital trails range from patterns in the way that students type, the frequency and duration of their interaction with online platforms (how often they log on, whether they do so in the middle of the night, and how long they spend between turning the pages of an uploaded document), and even the sorts of facial expressions they make when looking at a screen. All of these ‘trails’ are supposed to indicate something significant about how students learn and (hence) tell us something about how we might help them learn ‘better’: It is now possible to gather large data sets that show correlations between these patterns and certain educational outcomes. It is likely that, as digitization technology improves, the trails that can be stored will grow larger.

Fundamentally, the principal use of big data is to make predictions based on the patterns that it reveals. In education, big data can indicate whether a student is likely to perform badly in an upcoming test, because test performance may be correlated with a certain use (or low use) of an online platform through which students access learning material. They might indicate which pieces of material students find easiest to read and which they struggle with, based on the time that students spend looking at each page. It may be possible to tell, based on digitized correlations, whether a student is behaving in ways that indicate a high probability of dropping out, so that action can be taken to help them. This is one way in which big data might indicate when a teacher intervention can make a difference, perhaps earlier than a teacher might work this out by relying on their own perspective and its traditional method of tracking a student’s performance through observation.

The above is, we admit, a fairly thin description of how big data work in education.¹⁴ But it suffices to raise a certain number of concerns. We’ll now lay out what we take to be principal concerns about what sort of impact big data might have on the character of education. As we shall note, these worries are cautionary and some may be quite defeasible. As such, they may not prove to be absolute constraints against the use of big data in education. Our point is mainly to stress that they ought not to be ignored, and that they represent elements of a moral burden of proof where the use of educational big data are concerned.

Entrenching assumptions behind measurement

The first worry is that educational policies will be driven by what the data can actually show, rather than by what we should want it to show. One insight about big data (in general) is that it exacerbates a problem that is also true of smaller data sets, namely that it only includes

¹³ Here we draw on some recent surveys, Siemens (2013) and Wang (2016).

¹⁴ For a longer list of examples, though still a bit thin, see Long & Siemens (2011).

information that is measurable. This is sometimes called ‘ontic occlusion’. Fixating on data that show what can be measured sometimes leads to a failure to remember that the information is at best a partial representation of what one wishes to know about. As the old joke goes, one ends up looking for one’s car keys on the wrong side of the street, just because the street lighting is better on that side. The worry about ontic occlusion has particular force with respect to education. Simply put, the data on which we rely in any domain are only as useful as the presuppositions that have led us to rely on them in the first place. If these presuppositions are flawed, then this will infect the use of any such data. For example, we may assume that when a student spends longer looking at a page of an assigned piece of reading, they must be struggling with it. But our ability to gain very precise data about the time that students spend reading a page does not vindicate this assumption, it might just reinforce it instead. It is likely that some students are indeed struggling when they take a long time to read material. But other students may simply spend more time because they prefer to read slowly, or because they are getting more out of the material. These are qualitative differences that are left out of quantitative information that typically proves easier to measure and thus shows up disproportionately as data sets grow larger and more dependent for their formation on powers of computation.

There is one way in which this concern might be pushed further. It is possible that our conception of academic merit, or ‘learning’ more generally is already infected with a one-sided, perhaps even unjust conception of what makes for an excellent student. This sort of worry has been influentially expressed elsewhere.¹⁵ Again, we can offer a disanalogy with healthcare here. We generally know cancer when we see it, and the badness of cancer doesn’t presuppose anything about which ‘type’ of person has any particular excellence, or represents a type that other persons should become. Generalizing, there is no particularly strong presence of merit at play when we try to measure health outcomes. But educational outcomes are different. Big data may tell us which kinds of educational delivery result in which kinds of outcomes. But if the desired outcomes are shaped by implicit endorsement of one sort of group identity over another, then there are grounds for potentially quite serious moral concern.

Digitisation threatens social integration

Traditionally, children and higher education students have received their education in the same physical space, namely, a school or a campus. Historically, this has been necessary in order to deliver education efficiently. Apart from anything else, students need teachers, who are scarce. Students must therefore share teachers, and teachers can only be in one place at a time. But the value of shared learning environments isn’t restricted to overcoming scarcity of resources. It is also a means of ensuring that children, and in some cases adults, get access to a diverse peer group on a face-to-face basis. This is important to ensure the development of interpersonal skills, and to breed the kind of familiarity that can prevent the emergence of antagonistic inter-group relationships, such as stereotyping and prejudice, that are a hallmark of social segregated societies.

The rise of digitization puts this under threat. This is because the storage of data in digital form means that teachers and students do not have to occupy the same space in order for information to pass between them. Teachers can observe student performance by viewing digital trails, and tailor their instruction accordingly. There is no reason, in principle, that such

¹⁵ See especially the work of Iris Marion Young (1990: Ch.7).

action at a distance cannot work well in terms of the pedagogical relationship between a teacher and a student. Indeed, instances of this teaching style have been around for years before the rise of big data, particularly the radio broadcasts long practiced in Australia for the sake of children living in isolated parts of the country.

Our worry here is not about the way in which ‘teaching at a distance’ severs anything of value between teachers and students, but more the way in which it may pull students apart from each other. Schooling, and indeed university attendance, is largely about training for citizenship in ways that require interaction with peers, as equals, who are from different social groups. If digitization displaces the sharing of space in the learning process, it may destroy one of the most valuable means through which societies manage to pursue social integration. It is a moral imperative that the use of big data in education does not allow this to happen.

Patterns presume that students are (or ought to be) roughly homogenous

As we have said, applications of big data typically aim to make predictions about some subject matter based on prior patterns about correlations between how that subject matter is treated and the outcomes that emerge. But the very idea of ‘subject matter’ encourages the view that the data in question apply to some homogenous body, rather than a set of distinct individuals that are only to some degree similar.

There is a danger that, when a dataset shows a strong *prior* correlation between a certain style of learning and a certain level or type of student performance, then it will be inferred that there is something ‘wrong’ with a student who does not ‘respond’ according to how the prior pattern predicts they should respond. But there is nothing necessarily wrong with a student who doesn’t develop in the manner that other students do. This is perhaps where the disanalogy between education and medicine is strongest. If my muscle cells do not respond to an anti-inflammatory drug in the way that the muscle cells of most other patients do, then there is a strong sense in which my muscle cells are defective. But the same is not true in education. Sadly, there is a history of not properly respecting the way in which some students are different without this being a bad thing. During the 20th century, schools still forced left handed children to convert to using their right hands. This was burdensome and pointless, but motivated by an idea of what was normal that rested on little more than an arbitrary statistical trend. If big data supplies educational policy makers with other statistics that were simply less easy to observe than hand usage, there’s no guarantee that they will be any less arbitrary. There is some burden of proof, we think, on anyone who wants to rely on data to show that what’s passed off as ‘improved learning’ is not simply a more subtle form of arbitrary conformity.

This worry may also draw on the consideration, already mentioned, that big data tend (like much other computed data) to be quantitative rather than qualitative. As such, they place a special authority on students that are statistically ‘normal’ and suggest that difference is, by definition, an abnormality. This is wholly at odds with the liberal conception of education and its commitment to the view, long part of the liberal tradition, that individual difference is valuable and not something to be automatically corrected or suppressed. If this concern is not taken seriously when responding to big data, it increases the danger of ending up with the problem expressed by John Stuart Mill at the top of this essay.

Oversimplification of causation in the learning process

An ongoing worry about relying on data in the first place is that it gives rise to the so-called ‘tyranny of evidence’ – an institutional culture in which demands are made to demonstrate that the service or good being supplied is adhering to standards tied to the evidence available. The rise of *big data* follows a prior rise of an ‘evidence based’ culture of measuring and designing institutional performance. Applied to education, this worry parallels our disanalogy between educational services and medical services. Broadly speaking, medicine is interventionist in the sense that professionals work to ‘fix’ certain problems in ways that can, in principle, be tracked by evidence. But education is, in the words of one critic of the evidence-based approach, not ‘causal’, in the sense that teachers work with students in ways that allow students to provide feedback that is itself an input into what the teacher does next.¹⁶ This contrasts with many (though not all) medical services.

This worry might have limited application, particularly if big data can incorporate ‘trails’ that show not just how students initially respond to the use of certain techniques and materials, but also show which responses from teachers are correlated with which further responses from students. In this way, big data may track the more complex causal ‘back and forth’ that critics believe to be missing from the broader evidence-based agenda. But the worry has force nonetheless, insofar as some work will need to be done in order to ensure that appropriate datasets are selected.

The crowding out of goals whose pursuit lacks data

Big data may make certain opportunities available where learning is concerned. But opportunities carry costs. Data trails track some things but not others. It is reasonable to believe that data might, in principle, tell us when students are responding well and responding badly to styles of teaching. Big data may empower us with an opportunity to make students better at converting teaching into learning. But the fact that the opportunity has been created does not mean we should take it, as we may be sacrificing other educational goals that datasets are not helping us pursue. If we gear education increasingly towards activities for which big data provides guidance, then we may pay large opportunity costs with respect to education’s other goals.

These goals include, broadly speaking, a host of ends not necessarily coextensive with the largely academic process of learning. These include civic education, time spent playing, and any other benefit brought about by not being assigned a task on which one is later tested in an exam-style setting. The over-taught students of Mill’s day were much like the student who, today, emerges from school exhibiting a sort of ‘learned helplessness’. This may not be the fault of any reliance on any sort of data or evidence, but the increased use of data may eat into whatever other aspects of education are necessary to make students good at life, not just good at exams.

To be fair, crowding-out is certainly not a problem confined to the use of big data. It is something of a perennial hazard when it comes to confronting the trade-offs internal to education. Ultimately, part of the lesson here is that ‘education’ is not really a single good, even though our everyday vocabulary often proceeds as if it is. Really, education is a cluster of goods, some more measurable than others. Very generally, educational practice requires some care to ensure that some of the goods of education do not get pursued at the expense of the others. One

¹⁶ Biesta (2013: 10).

problem unrelated to big data is the tendency of academic results to dominate over whatever elements of education do not result in (and are not meant to result in) higher exam grades and test scores.¹⁷ The tendency of education to prioritize academic results probably causes (among other things) the sort of learned helplessness that frustrates higher education providers and employers, and anyone else invested in students being able to do things autonomously.

Education is coercive

Big data in healthcare, retail, etc. is being used so that providers of the relevant goods and services are able to make certain offers, and give certain advice, to consumers and clients. Big data may be used to tell someone that they face an unusually high risk of heart failure. Or it may be used to decide that a can of beans cost you \$4.21 at your local store even though it may cost someone else a different amount. Whatever objections we might have to this role of data in shaping our lives, the point is that these uses do not force us to do very much. We can ignore medical advice and adjust our consumption of beans. Education, at least in the K-12 range, is not like this.

The coercive nature of education does not count against any specific use of data, or any other evidence, in directing the course of education further. Instead, its point is that there simply needs to be an element of ‘hands off’ when it comes to the style of schooling, just to compensate for the fact that children have already been forced to show up to school every day irrespective of whether they feel like it, and typically without consultation as to which sort of school they might attend. Coercion in K-12 education is unavoidable and ultimately quite justifiable, but it inflates the moral cost of restricting children’s freedom once they are in school. It is not necessary that big data actually work to reduce freedom further, rather than just enhance the effectiveness of whatever tasks children are told to perform. Again our worry is cautionary: There are *special* reasons in education to not make the process ‘over designed’, for the sake of giving children just a bit of autonomy, both for its own sake and for what role it might have in preparing them for adulthood. In short, the outcomes of education are not all that matters. The process, as something that a child experiences at the time, matters too.

4. What are big data good for?

We have not tried to argue that the use of big data is necessarily at odds with the core goals of liberal education. Rather, our concern is that naïve use of big data may rely on an overly narrow view of what education is for, and which is naïve about certain hazards associated with unreflective reliance on what patterns can be revealed to us by large datasets.

There may yet be an important and valuable role for big data to play. It is uncontroversial, for example, to use big data to detect and monitor various inequalities in educational provision and outcomes. In addition, even if, as we have argued, there problems with relying on data to *enhance* the learning process, so far as the liberal conception of education is concerned, big data may be good at detecting distortion or disruption of the educational process as opposed to advancing that process directly. Important examples of distortion are plagiarism and cheating, which are a substantial problem in higher education, where there is an emphasis on assessment by way of the submission of long pieces of prose work. Datasets on work that has already been submitted help detect when this work has been

¹⁷ See for example one of the author’s discussion of the educational arms race in Halliday (2016).

duplicated via plagiarism. It is, of course, harder to detect so called ‘contract cheating’, when a student pays a third party to write an essay that is not already stored in a data set (because it has never been submitted before). But the rise of learning analytics has led to the development of software that can authenticate students by way of their typing styles. This provides some potential to identify when a student has typed the essay that they in fact submit.¹⁸ Generalizing, such cases provide evidence that algorithms processing big data do not need to have a homogenizing effect on students. Indeed, reduction of cheating is an important part of ensuring that students develop their individuality, simply by making sure that they do the assigned work themselves.

Most of the worries expressed in the preceding section have force in the K-12 range that they may not have in adult education. For one thing, adult education is typically not coercive, as adults seek it out voluntarily. It may also be under less pressure to deliver social integration, as unlike schooling, a relative minority of people undertake it. And whereas K-12 education has quite general and comprehensive ambitions, adult education is often narrowly vocational, and its users are often seeking to pick up a very specific set of skills to the exclusion of much else. For these reasons, big data might enhance adult education in ways less beset by hazards.

Returning to the Millian perspective, it is worth suggesting that some use of big data may, after all, count as an experiment in itself. At least, it would be anti-Millian, and thus in one way quite illiberal, to simply dismiss big data without trying it. True, we should note that some experiments reach a stage where it becomes pointless to continue repeating them. There is no longer much to be learned by allowing people to smoke dozens of cigarettes each day or to ride motorcycles without helmets. We know what happens when people do this, and these cases are among those where some restriction on freedom might not run afoul of a Millian concern to permit experiments in living. (There remains the question of whether such activities nevertheless enhance an individual’s quality of life given certain conceptions of the good, but that’s a separate concern from the idea that the rest of society has something to learn from seeing what happens.) But if there is something to learn from big data in education, any suspicion that big data might be damaging might be put to the test rather than allowed to pass as evidence in itself – as long as it is remembered that testing is what’s being done and that we should make every effort to measure the results properly, wary of such things as causal feedback loops, confirmation bias, and so on.

Ultimately, our concern is not really with big data *per se* but with philosophically naïve endorsement of its use to change the way in which education works. We want to signal that the liberal conception of education may find some space for big data, much as it might for ‘small data’ when used in similar ways. The important point is that data might help create an environment relatively free from certain obstructions to what may remain an open, somewhat unplanned, and experimental educational process. This is a far cry from any scenario in which increasing amounts of data are used to predict and plan educational outcomes from start to finish. Education is among the most important goods that people consume. But this does not change the fact that there are hazards associated with what might otherwise appear as irresistible opportunities to improve its delivery.

¹⁸ Software development to identify contract cheating (as opposed to mere plagiarism) is still in its early developmental phases. For a discussion of what’s at stake where cheating is concerned, see Walker & Townley (2012).

Acknowledgments

The authors wish to thank participants at the conference *Big Data and Education: Ethical and Moral Challenges*, held at the University of Haifa in May 2017. In addition, the authors thank an anonymous referee for comments on a subsequent manuscript.

Funding declaration

Daniel Halliday's work on this paper was supported by a grant from the Center for Ethics and Education, as part of their program supporting philosophy as it relates to educational policy and practice. He gratefully acknowledges the Center's support.

REFERENCES

- Anderson, E.
—2007 'Fair Equality of Opportunity in Education: A Democratic Equality Perspective' *Ethics* 117(4): 595-622.
- Biesta, G.
—2007 'Why 'What Works' Won't Work: Evidence-Based Practice and the Democratic Deficit in Educational Research' *Educational Theory* 57(1): 1-22.
- Brighouse, H.
—2000 *School Choice and Social Justice* (Oxford: Oxford University Press).
—2006 *On Education* (London: Routledge).
- Brighouse, H. & Swift, A.
—2014 *Family Values: The Ethics of Parent-Child Relationships* (Princeton, NJ: Princeton University Press).
- Callan, E.
—1997 *Creating Citizens: Political Education and Liberal Democracy* (New York, NY: Oxford University Press).
- Clayton, M.
—2006 *Justice and Legitimacy in Upbringing* (Oxford: Oxford University Press).
—forthcoming *Independence for Children* (Oxford: Oxford University Press).
- Deresiewicz, W.
—2015 *Excellent Sheep: The Miseducation of the American Elite and the Way to a Meaningful Life* (New York, NY: Free Press).
- Feinberg, J.
—1992 'The Child's Right to an Open Future' in Feinberg, *Freedom and Fulfillment: Philosophical Essays*. (Princeton, NJ: Princeton University Press).
- Gutmann, A.
—1999 *Democratic Education* 2nd ed. (Princeton, NJ: Princeton University Press).
- Halliday, D.
—2016 'Private Education, Positional Goods, and the Arms Race Problem' *Politics, Philosophy & Economics* 15(2): 150-169
- Levinson, M.
—1999 *The Demands of Liberal Education* (Oxford: Oxford University Press).
- Long, P. & Siemens, G.
—2011 'Penetrating the Fog: Analytics in Learning and Education' *EDUCAUSE Review* 46(5): 31-40.
- MacFarlane, K.
—forthcoming 'Education, Sufficiency, and the Relational Egalitarian Ideal' *Journal of Applied Philosophy*

- MacMullen, I.
 —2007 *Faith in Schools? Autonomy, Citizenship, and Religious Education in the Liberal State* (Princeton, NJ: Princeton University Press).
- Mayer-Schoenberger, V., & Cukier, K.
 —2013 *Big Data* (New York, NY: Houghton-Mifflin).
- Mill, J.S.
 —1989a *Autobiography* (London: Penguin Books).
 —1989b *On Liberty and Other Writings*, ed. S Collini (New York, NY: Cambridge University Press).
- Mittelstadt, B. D., & Floridi, L.
 —2015 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' *Science and Engineering Ethics* 22(2): 303-341.
- O'Neill, C.
 —2016 *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York, NY: Crown Publishing).
- Rawls, J.
 —1993 *Political Liberalism* expanded edition (New York, NY: Columbia University Press).
- Reich, R.
 —2002 *Bridging Liberalism and Multiculturalism in American Education* (Chicago: Chicago University Press).
- Satz, D.
 —2007 'Equality, Adequacy, and Education for Citizenship' *Ethics* 117(4): 623-648.
- Siemens, G.
 —2013 'Learning Analytics: The Emergence of a Discipline' *American Behavioral Scientist* 57(10): 1380-1400.
- Slade, S. & Prinsloo, P.
 —2013 'Learning Analytics: Ethical Issues and Dilemmas' *American Behavioral Scientist* 57(10): 1510-1529.
- Smith, A.
 —1993 *An Inquiry into the Nature and Causes of the Wealth of Nations* (Indianapolis, IN: Hackett).
- Walker, M. & Townley, C.
 —2012 'Contract Cheating: A New Challenge for Academic Honesty?' *Journal of Academic Ethics* 10(1): 27-44
- Wang, Y.
 —2016 'Big Opportunities and Big Concerns of Big Data in Education' *Tech Trends* 60(4): 381-384.
- Young, I.
 —1990 *Justice and The Politics of Difference* (Princeton, NJ: Princeton University Press).

Author Biography

Matthew Clayton is professor of political theory at the University of Warwick. He works on distributive justice, the foundations of liberal political thought, and also certain kinds of applied philosophy, such as the philosophy of education. His publications include *Justice and Legitimacy in Upbringing* (Oxford University Press, 2006) and two co-edited books: *The Ideal of Equality* (Palgrave, 2002), and *Social Justice* (Blackwell, 2004).

Daniel Halliday is senior lecturer in philosophy at the University of Melbourne. His current research focuses on problems of justice and political economy. His book *The Inheritance of Wealth: Justice, Equality and the Right to Bequeath* will be published by Oxford University press in 2018.

Contact:

Matthew Clayton
Politics and International
Studies, Social Sciences Building
Warwick University
Coventry, CV4 7AL, United Kingdom
m.g.clayton@warwick.ac.uk

Daniel Halliday
Philosophy, Arts West Building
University of Melbourne
VIC 3010, Australia
daniel.halliday@unimelb.edu.au